Calculation of free energy landscapes: A Histogram Reweighted Metadynamics approach

Jens Smiatek* and Andreas Heuer†

Institut für Physikalische Chemie, Westfälische Wilhelms-Universität Münster, D-48149 Münster, Germany

(Dated: March 3, 2011)

We present an efficient method for the calculation of free energy landscapes. Our approach involves a history dependent bias potential which is evaluated on a grid. The corresponding free energy landscape is constructed via a histogram reweighting procedure a posteriori. Due to the presence of the bias potential, it can be also used to accelerate rare events. In addition, the calculated free energy landscape is not restricted to the actual choice of collective variables and can in principle be extended to auxiliary variables of interest without further numerical effort. The applicability is shown for several examples. We present numerical results for the alanine dipeptide and the Met-Enkephalin in explicit solution to illustrate our approach. Furthermore we derive an empirical formula that allows the prediction of the computational cost for the ordinary metadynamics variant in comparison to our approach which is validated by a dimensionless representation.

Keywords: Free energy landscape, Essential dynamics, Metadynamics variants, Histogram reweighting

INTRODUCTION

The characteristics and the behaviour of physical systems can be understood and predicted by the investigation of the underlying free energy landscape. The knowledge of this often complex surface allows one to determine reaction pathways for chemical reactions as well as stable configurations for proteins [1–5, 9] and glass-forming systems [6–8]. A successful tool for these studies are computer simulations which help to explore the characteristics of the system in detail. Naturally, during the last years a lot of effort has been spent to develop novel efficient and time-saving methods.

Specifically in protein sciences, the free energy approach is promising for the investigation of unknown folding mechanisms. To achieve a proper description of the dynamical behaviour, the trajectories may be projected onto a set of well chosen collective variables [1, 2] which allow to define an effective energy landscape. These variables can be interpreted as effective reaction coordinates spanning the phase space respectively the free energy space of the system. Well known collective variables are the number of native contacts, dihedral angles and the essential eigenvectors or principal components of the protein [1, 2]. Thus a set of well suited reaction coordinates offers the opportunity to describe the folding or unfolding of a protein adequately without too much loss of information.

However, it has to be mentioned that in some cases suitable collective variables are hard to determine. The investigation of the underlying dynamics of the system can then be performed by computational methods like discrete path sampling [8, 10] and transition path sampling [11, 12] to overcome this problem. These techniques represent computational algorithms which are not dependent on the usage of collective variables.

Experiments and computer simulations have shown that naturally occuring proteins often have free energy landscapes with global funneling properties [1–3, 5, 8, 9]. Nevertheless local minima can occur which represent trapped conformations of the protein [1, 9]. The transitions between these minima which are called rare events occur on timescales which are typically not accessible in computer simulations. Over the last years several numerical methods have been developed to accelarate rare events and to compute the underlying free energy landscapes. In the following we will mention the most common ones which have been well established in the scientific community.

Often used methods are thermodynamic integration [13–15] and free energy perturbation [16]. These techniques are well suited to calculate static properties like hydration energies but normally fail in the calculation of free energy differences between specific folding mechanisms. To overcome this situation and to gain insight into dynamical quantities, more sophisticated ideas have to be adopted.

The umbrella sampling method [17] allows to determine free energy differences via the corresponding probability distributions of the whole accessible phase space. An effective additional biasing potential drives the system to more unlikely regions such that the whole phase space is visited. This idea has also been used in adaptive force biased calculations [18, 19], steered as well as adiabatic Molecular Dynamics simulations [20, 21] and multicanonical sampling approaches [22–24]. Another promising method for the investigation of folding properties is the replica-exchange algorithm [25].

Specifically in the last years several techniques have been proposed which employ an adaptively varying bias potential as an estimate for the free energy. Examples for these methods are the Metadynamics algorithm in its several variants [26, 27] and the related local elevation method respectively the conformational flooding approach [28, 29].

The main idea of the Metadynamics method relies on a successive flattening of the free energy landscape by an additional potential energy in form of small gaussian hills which are positioned at relaxed times at the present location of the system. Then the free energy can be estimated as a negative mirror of the potential energy which is exact in the limit of long simulation times [27]. It has been shown that this method is in principle valid for all collective variables and offers a broad range of applicability [27]. Even the calculation of microscopic averages within the metadynamics scheme has been successfully examined [30]. Nevertheless the drastic error dependence on the parameters of the algorithm [31, 32] in the original metadynamics scheme has to be mentioned. Keeping the errors within tolerable ranges results in a drastic increase of simulation time. To avoid local bumps in the landscape and to keep the errors low, extensions of the existing Metadynamics method [33–36] have been proposed which offer a more reliable derivation.

In this paper we present a Histogram reweighted Metadynamics technique which offers the opportunity to overcome some drawbacks of the original method and its variants. Our method is grid-based which means that the energy landscape spanned by the collective variables is divided into several regions. By adding a short range cut-off biasing potential which is only evaluated at the grid points of a predefined grid, the corresponding simulation time can be massively decreased. Finally, the corresponding biased probability distributions are reweighted by histogram techniques [37–40] to compute the free energy landscape a posteriori. It has been noticed that these techniques offer a low error dependent description [39, 40].

Additionally a projection scheme is introduced which allows the investigation of further collective variables if the probability distributions of the studied variables are well overlapping and if the projected coordinates are well-suited and fastly varying. Although it is clearly impossible to reconstruct the whole highly dimensional phase space of the protein, our projection scheme allows to investigate additional low dimensional reaction coordinates like dihedral angles or spatial distributions in good agreement to directly derived energy landscapes. As a consequence of this scheme, no further simulation time is required as the analysis can be performed a posteriori and the landscape is not restricted to the actual chosen set of coordinates.

This becomes useful by regarding the fact that difficulties for predefined collective variables may arise, if they are not the true reaction coordinates of the system [27, 41] or give unsuitable descriptions of the corresponding free energy landscapes. This is in particular a problem if the underlying folding or unfolding mechanisms are too complex for being projected onto a low dimensional subspace which may lead to wrong conclusions [27] or if the phase space cannot be sampled efficiently. In these cases,

discrete path sampling [8, 10], transition path sampling [11, 12] and additional methods [41–44] allow to investigate the hidden complexity of the free energy landscape without the usage of collective variables as alternative approaches.

By comparison to other approaches, it has to be noticed that in recent publications [32, 45, 46] the usage of umbrella sampled biased probability distributions respectively using histogram reweighting procedures [48] for the correction of free energy landscapes has also been claimed. Even the evaluation of the underlying potential on a grid point has been recently proposed for metadynamics [46] as well as for a closely related adaptive bias molecular dynamics scheme [47]. Both ingredients have also been applied in a grid-based adaptive umbrella sampling scheme [49] over a decade ago and the above mentioned techniques are implemented in common software packages [50–55].

A main reason for the introduction of grid evaluations is a massive decrease in computation time which scales by a constant factor. This fact is also used in [46] where smoothly truncated Gaussians are evaluated by a kd-tree. Further realizations of grid techniques incorporate adaptive grids where the potential energy is calculated by a polynomial extrapolation on the grid points [51, 53], choosing the potential of a close grid point as the biasing potential on the particle [51] or applying cut-off radii for the gaussians [51]. It was also shown that the above mentioned adaptive biased grid approach in its computational cost scales linearly with simulation time in contrast to ordinary metadynamics [47]. Here cubic B-splines are used for the evaluation of the biasing potential on the corresponding grid points.

Our approach has the advantage of a well-defined simple potential all over the energy surface. This allows to change the resolution of the grid on-the-fly even after the simulations have finished to resolve regions of specific interest in more detail. In addition, it is easy to implement and scales with a constant number $\mathcal{O}(2d)$ of calculations per timestep, where d is the number of collective variables, respectively the dimension of the grid.

As a further remark, our potential energy landscape is purely used as a biasing potential to accelerate rare events. Hence, the fine resolution of the free energy landscape is achieved by histogram techniques. This allows us to use a very rough potential energy surface created in a short simulation time. Additionally the bias force is always well-defined such that discontinuities are avoided. Therefore our method in its interpretation is closely related to the Wang-Landau approach [56] and adaptive umbrella biasing techniques [49] with the effectiveness of [47]. This finally results in a simple and robust algorithm which is easy to implement in common public software packages like GROMACS [58–60] or other programs and allows to tune the resolution of the landscape on demand easily after the simulations have finished.

We further derived an empirical formula for the computational cost required for the ordinary metadynamics scheme in comparison to our approach. Thus we were able to show that the grid technique for reasons of computational efficiency is often more preferable especially for a large number of hills in comparison to the number of atoms in the system.

The paper is organized as follows. In the second section we introduce the method and the theoretical background. The third and the fourth section illustrate model test cases and the numerical details of the following peptide simulations. The fifth section presents the results for the alanine dipeptide and the Met-Enkephalin. In the sixth section the values for the computational cost required for both methods are investigated and an empirical formula is derived. We conclude with a brief summary in the last section.

HISTOGRAM REWEIGHTED METADYNAMICS: THEORETICAL BACKGROUND

The system we consider is described by a set of coordinates x evolving under the action of dynamics following the trajectory $x(t)$ and described by a canonical equilibrium distribution at temperature T. The set of coordinates x may include atomic positions or angles as well as any other auxiliary collective variable representing the characteristics of the system.

If the system shows metastability, some regions separated by large energy barriers cannot be explored by the evolution of the trajectories in a reasonable simulation time. Guided by the conventional metadynamics approach an additional potential has to be added at specific constant times $t_1, t_2, ... t_N$ on the trajectory $x(t)$ to overcome the barriers and to accelerate the rare events. It has to be ensured that the protein relaxes between these times such that the system diffuses in the next local minimum. For the force evaluation in the ordinary Metadynamics method, the sum over all previously added gaussian functions has to be performed [26] which results in a strong increase of computation time. To avoid this increase, we use an approach where the additional potential energy is evaluated on the grid points of a predefined grid closely related to [46, 47, 49], spanning the whole range of the accessible phase space in the set of collective variables x. The grid points $x_{G,i}$ are separated by the grid constant σ_G, which is the distance between two neighboring grid points in one dimension. The system now evolves in time moving over the energy landscape covered by the grid. At the times $t = t_1, t_2, ... t_N$ the following potential energy

$$\Delta V_T(x_{G,i}, t) = \mathcal{G}(x - x_{G,i}) \, \omega \qquad (1)$$

with

$$\mathcal{G}(x - x_{G,i}) = e^{-\frac{(x(t) - x_{G,i})^2}{2x_c^2}} \left(1 - \left(\frac{|x(t) - x_{G,i}|}{x_c}\right)\right) \qquad (2)$$

defined by

$$\begin{aligned} \mathcal{G}(x - x_{G,i}) \neq 0 &: |x(t) - x_{G,i}| < x_c \\ \mathcal{G}(x - x_{G,i}) = 0 &: |x(t) - x_{G,i}| \geq x_c \end{aligned} \qquad (3)$$

is evaluated on all grid points N_{GK} whose distance from $x(t)$ is closer than the cut-off radius x_c. The magnitude of the potential energy is given by the height ω like in the conventional metadynamics scheme. As it has been discussed in [27] the values for ω have to be low.

The potential energy at the grid points then evolves in simulation time with

$$V_T(x_{G,i}, t_N) = V_T(x_{G,i}, t_{N-1}) + \Delta V_T(x_{G,i}, t_N) \qquad (4)$$

emerging rapidly if the specific neighborhood of the grid point is often visited by the trajectory, e.g. in free energy minima. Then we need to update $2d$ values of grid points for each calculation, where d is the number of collective variables or in other words the dimension of the grid. The biasing potential exerted on the system at times $t \geq t_N$ yields

$$V_B(x, t) = \sum_i^{N_{GK}} \mathcal{G}(x - x_{G,i}) \, V_T(x_{G,i}, t_N) \qquad (5)$$

where the summation is over the number of grid points N_{GK} within x_c, with the actual value of the potential $V_T(x_{G,i}, t_N)$ of each grid point at $x_{G,i}$ as defined in Eqn. (4) and the weighting factor $\mathcal{G}(x - x_{G,i})$ of Eqn. (2). The resulting local minimum between two grid points can usually be neglected if the distance between neighboring grid points σ_G is small compared to the typical expected diameter of a free energy minimum, respectively a high resolution of the grid. Repeating the whole scheme allows to fill the minima efficiently until the potential energy is reached to overcome the energetic barriers. The energy landscape with constant values $V_T(x_{G,i})$ and the corresponding continuous function $V_B(x)$ are finally used in the biased simulation runs to produce flat probability distributions. An illustration of the scheme is presented in Fig. 1.

It has to be noticed that the specific choice of the biasing potential is arbitrary [27, 46] although in combination with the grid potential evaluation it can be pointed out that the function proposed in Eqn. (2) offers the advantage of a fast computation. Due to the well-defined potential at every position, it can be even used to achieve a finer resolution of the grid in combination with histogram reweighting techniques. As a further remark, it has to be mentioned that the final biased simulations as well as the previous simulation runs for the construction of the biasing potential can be trivially parallelized in the spirit of the multiple walker metadynamics technique [61] to save computation time.

In the following we give a short description of the histogram reweighting technique. The free energy landscape
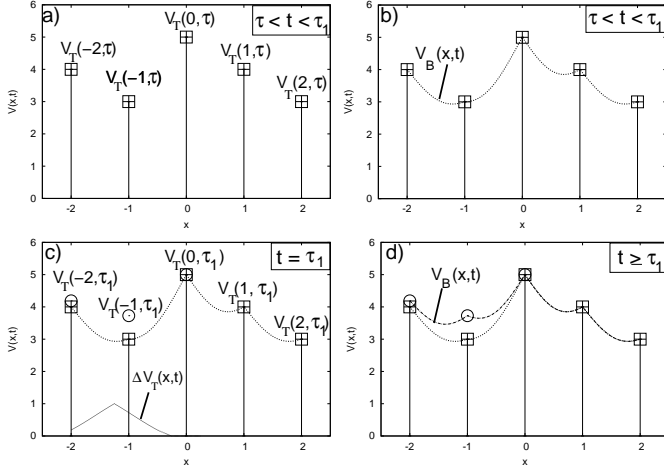
FIG. 1: Illustration of the Histogram Reweighted Metadynamics scheme with $x_c = \sigma_G = 1\sigma$. Panel A shows the constant potential energy $V_T(x_{G,i}, \tau)$ of the grid points at times $\tau < t < \tau_1$. The biasing potential energy $V_B(x,t)$ of Eqn. (5) for times $\tau < t < \tau_1$ is shown in Panel B. At times $t = \tau_1$ (Panel C) a new gaussian function $\Delta V_T(x,t)$ is placed at the actual position of the system at $x = -1.25\sigma$ which is only evaluated at the grid points such that $\Delta V_T(x_{G,i})$ is summed to the old potential $V_T(x_{G,i}, \tau)$ via Eqn. (4) yielding the new values $V_T(x_{G,i}, \tau_1)$(shown as circles). The resulting new biasing potential function $V_B(x,t)$ is shown in Panel D.

is finally calculated via the WHAM equations [37–40] by a reweighting of the biased probability distributions. This procedure allows the calculation of the free energy within a finite tolerance value [40] and its combination to the biasing potential for refinement and construction has been shown to be advantageous [27, 31, 46, 57]. The scheme is given by two equations

$$P(x) = \frac{\sum_{i=1}^{N_{sim}} n_i(x)}{\sum_{i=1}^{N_{sim}} N_i e^{-(F_i - V_B(x))/k_B T}} \quad (6)$$

and

$$F_i = -k_B T \log\left(\sum_{x_{bins}} P(x) e^{-V_B(x)/k_B T}\right) \quad (7)$$

which allow to compute the best estimate of the unbiased probability $P(x)$ where N_{sim} denotes the number of independent simulations, $n_i(x)$ the number of counts in a histogram bin associated with x, $V_B(x)$ is the final constant biasing potential of Eqn. (5) at position x and the free energy shift is F_i for each simulation with thermal energy $k_B T$. Due to the fact that F_i and $P(x)$ are unknowns, Eqn. (6) and Eqn. (7) have to be solved by iteration to self-consistency within a predefined tolerance value. The values of $P(x)$ can then be used to calculate the resulting free energy of the bin via

$$F(x) = -RT \log\left(\frac{P(x)}{P(0)}\right) \quad (8)$$

where $P(0)$ is a reference value and R denotes the molar gas constant [39]. It has to be pointed out, that the error in the energy is strongly dependent on the resolution of the histogram leading to more or less pronounced approximations. For a detailed description of the method we refer the reader to Refs. [38, 39].

As it was mentioned in the introduction, the main part of a successful calculation relies on the choice of appropriate collective variables on which the free energy surface is spanned. Well-suited collective variables are dihedral angles as well as the essential eigenvectors of the system [62]. It has been recently demonstrated that the application of eigenvectors as collective variables results in adequate descriptions of the folding mechanisms and of the free energy landscape in ordinary Metadynamics computations [63, 64]. Thus we follow this approach due to the assumption that nearly all relevant motion is captured in the first eigenvectors of the system [62]. This is remarkable by regarding the fact that drastic difficulties appear if wrong reaction coordinates are chosen which do not capture the main motion [27, 41]. But nevertheless, it has again to be remarked that the usage of eigenvectors is not a guarantee for correctness of the free energy landscape due to hidden complexities in higher dimensional collective variables [27]. In the following we give a brief description of the method.

The essential eigenvectors can be calculated by the superimposed coordinates $\vec{r}$ of N atoms of the system which build the covariance matrix $\mathbf{C}$ via

$$\mathbf{C} = <(\vec{r} - <\vec{r}>)(\vec{r} - <\vec{r}>)^T> \quad (9)$$

where $i, j = 1, 2, \ldots 3N$ and the time-averaged or reference value is denoted by the brackets $< \ldots >$. The diagonalization of $\mathbf{C}$ leads to

$$\mathbf{C} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^{-1} \quad (10)$$

where $\mathbf{E}$ is a matrix of eigenvectors and $\mathbf{\Lambda}$ is a matrix of eigenvalues marking the positional fluctuations. Sorting the eigenvalues in decreasing order allows to identify the largest positional fluctuations with all important structural transitions by the first eigenvectors which form the essential dynamics of the system [62]. The projection at time t on the i−th eigenvector is then defined by

$$p_i(t) = (\vec{r}(t) - <\vec{r}>) \cdot \vec{e}_i \quad (11)$$

with the specific eigenvectors ranging from $i = 1, 2 \ldots 3N$.

If the motion of the system is not restricted to the essential subspace in the biased simulations such that even lower eigenvectors contribute, the potential biasing energy of Eqn. (5) even activates the motion of the remaining subspace. Hence the concerted motion of all eigenvectors is influenced by the biasing energy. Therefore a projection scheme allows to construct the free energy landscapes of additional collective variables like lower eigenvectors or dihedral angles a posteriori.

The main idea is that the applied bias potential energy drives the system through structural transitions which could be in principle also observed in additional collective variables. The occurrence of the variables $x(t)$ have to be projected onto the new collective variable $\xi(t)$ exposed at the same time. Formally, this is given by

$$f \; : \; x(t) \rightarrow \xi(t), \tag{12}$$

together with the corresponding biasing potential energy

$$f \; : \; V_B(x,t) \rightarrow V_B(\xi,t). \tag{13}$$

After the final timestep τ, the constant potential energy at the new grid points $\xi_{G,i}$ can be evaluated by using the maximum last values of $V_B(\xi,\tau)$. Choosing the maximum value for $V_B(\xi_{G,i})$ within certain small regions δ, where δ is the half distance between two grid points gives

$$V_B(\xi_{G,i}) = \max(V_B(\xi_{G,i} \pm \delta, \tau)) \tag{14}$$

for the construction of the potential energy landscape in the new set of chosen reaction coordinates. The values for $V_B(\xi_{G,i})$ can now be inserted into the Eqns. (6) and (7) with the corresponding biased probabilities $P(\xi_{G,i})$. The final potential energy landscape in the new set of collective variables of Eqn. (14) is evaluated by the existing data and by explicit calculation of each $\xi(t)$. Sampling the biased probabilities in the new set of variables finally allows to determine the unbiased probabilities by using the histogram reweighting procedure in the new set of collective variables at any resolution.

Again it has to be ensured that the motion is not constrained and that the potential energy even activates lower eigenvectors to establish a free system behaviour. In principle all collective variables which are fastly varying and cover a small subspace can be seen as suitable reaction coordinates if a sufficient simulation time is given. Additionally they also have to clearly distinguish between different states of conformations, can be well sampled meaning a good overlap in the distribution functions and covering of the phase space and show no hysteresis effects due to hidden complexities [27]. Several publications discuss this problem and new techniques to overcome it in more detail [8, 10–12, 41–44]. If these requirements are fulfilled, the projected energy landscape into the new collective variables can be seen as reliable. Nevertheless, it has to be noticed that the validity of a landscape is related to the suitability of the chosen projected collective variables. This is also true for the original chosen coordinates.

1- AND 2-DIMENSIONAL MODEL POTENTIAL

The first test case for our approach is a particle confined in a one-dimensional well defined potential

$$V(x) = \epsilon \left[\left(\frac{x}{\sigma} \right)^4 - 10 \left(\frac{x}{\sigma} \right)^2 \right] \tag{15}$$

where ϵ is the unit thermal energy $k_B T$ and the length unit is given by σ. We performed a Monte Carlo simulation with the Metropolis criterion [65] consisting of 10^6 timesteps where every $500-$th step a potential energy with height $\omega = 0.25\epsilon$ has been set. The grid points were separated by a grid constant $\sigma_G = 0.1\sigma$. The cutoff radius for the construction of potential function has been chosen to $x_c = 0.2\sigma = 2\sigma_G$ and $x_c = \sigma_G$ for the evaluation of the biasing potential. Having constructed the biasing energy landscape, we performed four simulations with 10^6 timesteps to derive the biased probability distribution functions which have been reweighted by the WHAM equations (Eqns. (6) and (7)). Fig. 2 presents the numerical results for the potential of Eqn. (15). The numerical values are in good agreement to the theoretical results of the test potential except for $|x| > 3.0\sigma$. The reason for this misbehaviour can be related to the low statistical accuracy of the probability distribution [66] which is caused by the rapid increase of the model potential at these points. The constant potential $V_B(x)$ of Eqn. (5) as well as the corresponding discrete potential energy values of the grid points of Eqn. (4), which have been used for the biased simulations are shown in the lower panel of Fig. 2. Nevertheless the results of this simple model potential have shown that our method works and produces accurate results in one dimension.

Another 2-dimensional test potential is given by

$$V(x,y) = \epsilon \left[\left(\frac{x}{\sigma} \right)^4 + \left(\frac{y}{\sigma} \right)^4 - 10 \left(\left(\frac{x}{\sigma} \right)^2 + \left(\frac{y}{\sigma} \right)^2 \right) + 50 \right] \tag{16}$$

which represents four energetic minima. We performed a couple of Langevin Dynamics simulations [67] obeying the Langevin equation which is given by

$$\vec{F}_i = -\zeta \vec{v}_i + \vec{\eta}_i \tag{17}$$

where the force $\vec{F}_i$ on a particle depends on the friction coefficient ζ, the velocity of the particle $\vec{v}_i$ and the stochastic force $\vec{\eta}_i$. The stochastic force represents the thermal brownian motion of the particle with the following properties

$$< \vec{\eta}_i(t) >= 0 \tag{18}$$

and

$$< \eta_{i\alpha}(t)\eta_{j\beta}(t') >= 2\zeta k_B T \delta_{ij} \delta_{\alpha\beta} \delta(t - t') \tag{19}$$

which ensures the presence of the fluctuation dissipation relation. Thus the stochastic force is delta-correlated white noise which ensures a canonical ensemble at thermal equilibrium. We chose $\zeta = 1\sqrt{m\epsilon}/\sigma$ with the mass m, the temperature $T = 1$ and ran a simulation of 10^7 timesteps, where every 1000-th step a gaussian function has been set. We chose as values for the parameters $\sigma_G = 0.25\sigma$, $x_c = 0.5\sigma$, $\omega = 0.1\epsilon$ and we used a timestep of $\delta t = 0.01\sigma/\sqrt{m\epsilon}$. The final values for $V_T(x_{G,i})$ of

FIG. 3: Values of $V_T(x_{G,i})$ which have been used in the biased simulations for the 2-dimensional test potential of Eqn. (16).
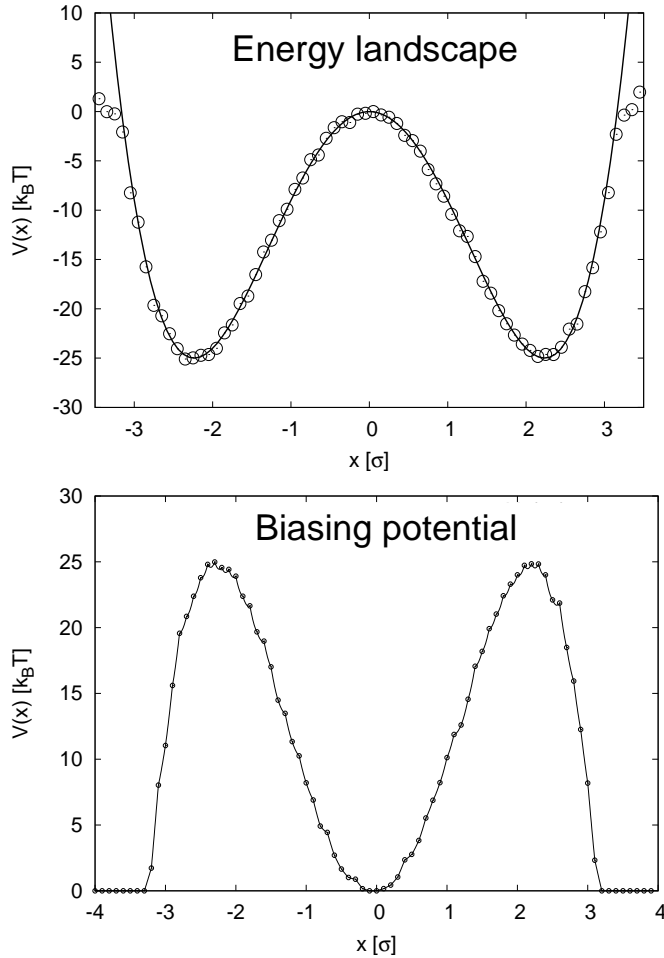
FIG. 2: **Top:** Shifted energy landscape for the test potential of Eqn. (15). The circles correspond to the values derived by the Histogram reweighted Metadynamincs approach whereas the black line corresponds to Eqn. (15). The data points are in good agreement to Eqn. (15) except for values $|x| > 3.0\sigma$. This is due to the error of the statistical accuracy of the probability distribution which is caused by the rapid increase of the test potential. **Bottom:** Continous biasing potential $V_B(x)$ (solid line) which is used for the calculation of the biasing probabilities and the corresponding values of the potential energy $V_T(x_{G,i})$ (points) evaluated at the grid points.

FIG. 4: **Top:** Energy profile of Eqn. (16). **Bottom:** Resulting reweighted energy profile. The solid lines correspond to energy differences of $5k_BT$.

Eqn. (4) are shown in Fig. 3. These values have been used for the evaluation of the biasing potential of Eqn. (5). We performed four simulation runs with $2.5 \cdot 10^6$ timesteps where again every 1000-th step the position of the particle has been stored for the sampling of the underlying free energy landscape. The resulting reweighted energy profile is presented in Fig. 4. The good agreement to the energy profile of Eqn. (16) is obvious. Larger deviations only occur at positions $x, y > 3.5\sigma$ above $60\ k_BT$ due to poor statistical accuracy. The standard deviation to the correct profile is around 5% per value.
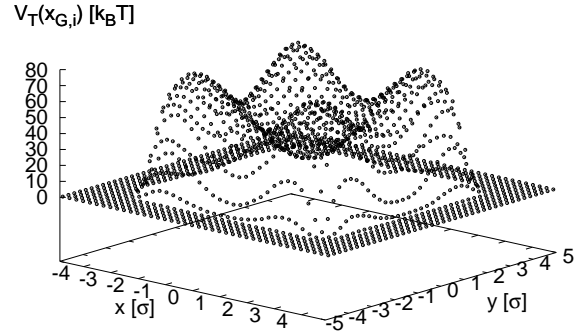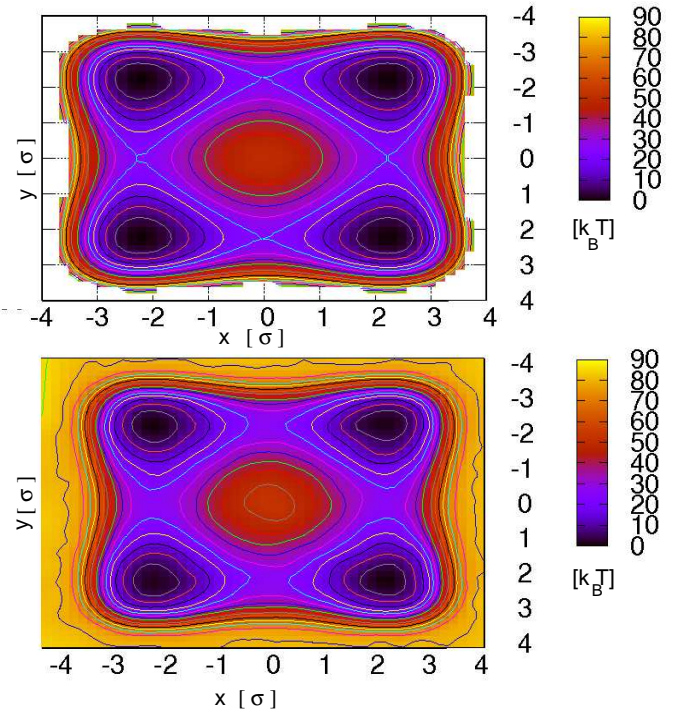
NUMERICAL DETAILS

All our Molecular Dynamics simulations have been performed by the software package GROMACS [58–60] in which our in-house written code for the Histogram reweighted Metadynamics method has been implemented. The source code is available on request by contacting one of the authors.
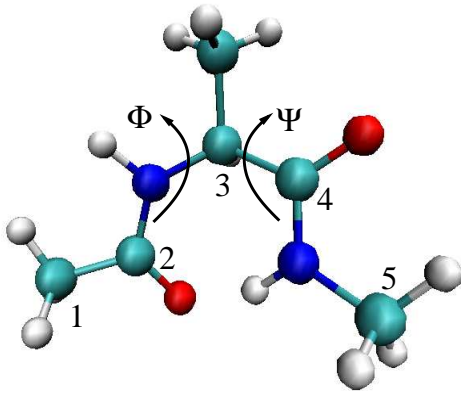
FIG. 5: Schematic representation of the alanine dipeptide with dihedral angles Φ and Ψ. The numbered C_α carbon atoms have been used for the eigenvector analysis.

Alanine dipeptide

We performed our Molecular Dynamics simulations by using the GROMACS ports of the AMBER94 force field [68, 69]. The simulation box contains 313 TIP3P water molecules [70] within a dimension of $2.005 \times 2.194 \times 2.192$ nm. The time step was 1 fs. The temperature $T = 300$ K was kept constant by a Nose-Hoover thermostat. Electrostatics have been calculated by the PME (Particle Mesh Ewald) method. All bonds have been constrained by the SHAKE algorithm [71].

After a short equilibration run of 350 ps we performed a 500 ps unbiased simulation to analyze the eigenvectors of the system constructed by the numbered C_α carbon atoms of Fig. 5. The corresponding potential energy landscape has been constructed in the phase space of the first two eigenvectors within a 2 ns simulation run. The grid constant has been chosen to $\sigma_G = 0.01$ nm, the height of the potential function has been set to $\omega = 0.2$ kJ/mol and the relaxation time was $\tau = 1$ ps. The cut-off radius x_c was 0.02 nm for hill setting and 0.01 nm for the evaluation of the biasing forces. The final biased probability distributions have been derived by four several independent 2 ns simulations with different temperatures $300, 310, 320$ and 330 K whose additional kinetic energy allows to explore the phase space more efficiently. Nevertheless the WHAM equations allow to combine the results for the different temperatures and give the results for a specific temperature if not too large deviations exist [39].

For a comparison of the eigenvector free energy landscape we further applied a conventional metadynamics scheme as described as in [63, 64]. The simulation time was 2 ns where every picosecond a gaussian hill of width 0.01 nm with height 0.2 kJ/mol has been set. Additionally we used the software plug-in PLUMED [50, 51] for a conventional metadynamics simulation to directly calcu-late the free energy landscape of the Ramachandran plot. All parameters are identical to the histogram reweighted metadynamics scheme.

Met-Enkephalin

The molecule consists of five residues TYR-GLY-GLY-PHE-MET and its molecular structure has been taken from the PDB entry 1PLW [72]. The force field was GROMOS96 [73]. The box size has been chosen to $3.214 \times 3.214 \times 3.214$ nm with 1058 SPC water molecules. The time step was 2 fs and the temperature was kept constant by a Nose-Hoover thermostat. As mentioned above, all bonds have been constrained by the SHAKE algorithm [71]. Electrostatics have been again calculated by the PME (Particle Mesh Ewald) method.

After a warm up phase of 350 ps we perfomed a 400 ps simulation run for the analysis of the corresponding eigenvectors of all atoms at temperature $T = 400$ K. We chose this high temperature to capture all necessary transitions in the essential first eigenvectors. The corresponding potential energy grid has been constructed in a 1 ns simulation run where the grid constant has been chosen to $\sigma_G = 0.01$ nm and the height of the potential function has been set to $\omega = 0.2$ kJ/mol with relaxation times of $\tau = 2$ ps. The cut-off radius x_c has been chosen to 0.2 nm respectively $x_c = 0.1$ nm as mentioned above. The biased simulations consist of three independent runs of $T = 300$ K and $T = 305$ K with 4 ns and $T = 310$ K with 1 ns simulation time. The corresponding timestep was $\delta t = 2$ fs. All energy landscapes have been reweighted for a temperature $T = 300$ K.

NUMERICAL RESULTS

Alanine Dipeptide

A well suited model system to test our approach is the alanine dipeptide (Fig. 5). Alanine dipeptide is one of the most studied model systems over the last years [44, 74–80]. The corresponding eigenvalues and the cumulative positional fluctuations are shown in Fig. 6. It is obvious that nearly 80% of the overall positional fluctuation can be described by the first two eigenvectors. Thus it can be assumed that all important structural transitions should be captured by using these variables as reaction coordinates [62].

The corresponding positions for the 300 K biased simulations are presented in the top (orange triangles) of Fig. 7 in contrast to an unperturbed simulation run (blue stars) for the same parameter sets. The bottom figure illustrates the values for the dihedral angles Φ and Ψ in the Ramachandran plot. It is obvious that the biasing potential energy landscape drives the system into more
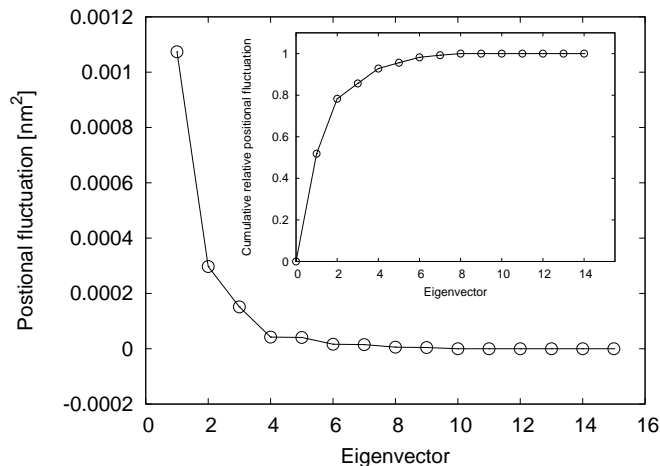
FIG. 6: Fluctuation and cumulative relative positional fluctuation (inset) of the eigenvectors. Nearly 80% of the overall motion of the alanine dipeptide can be described by the first two eigenvectors.
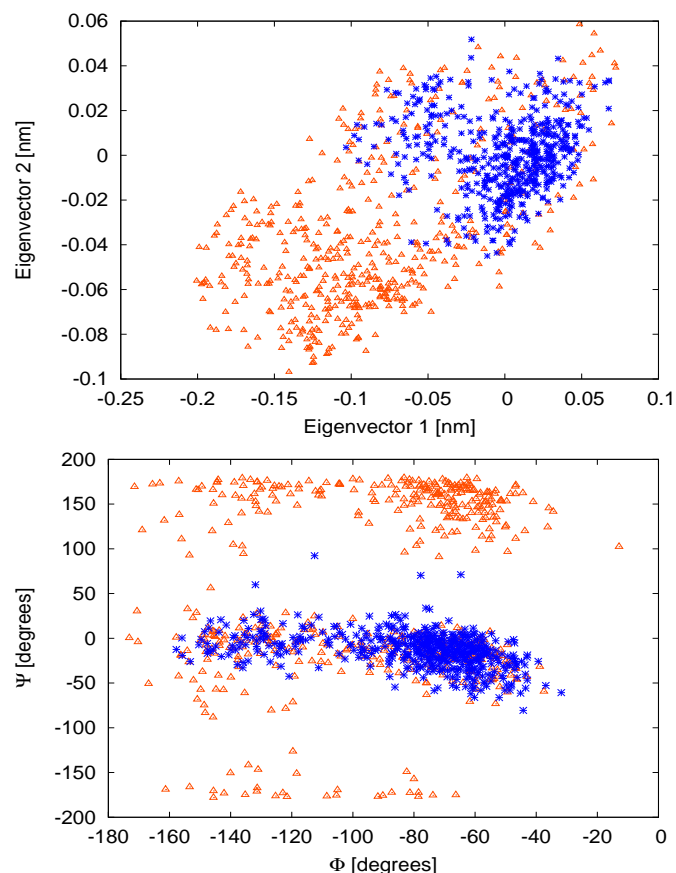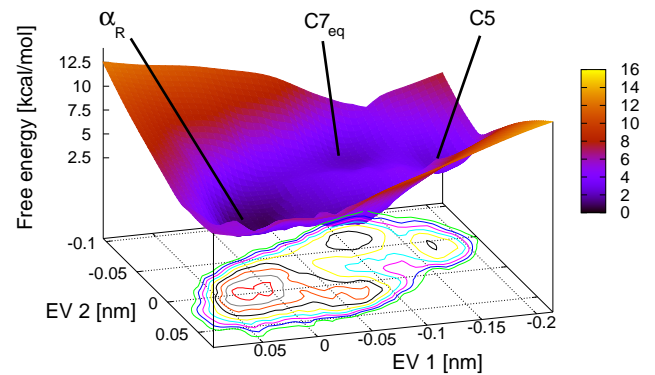

FIG. 8: Free energy landscape for the eigenvectors 1 and 2 of the alanine dipeptide. The lines correspond to energy differences of 0.5 kcal/mol.
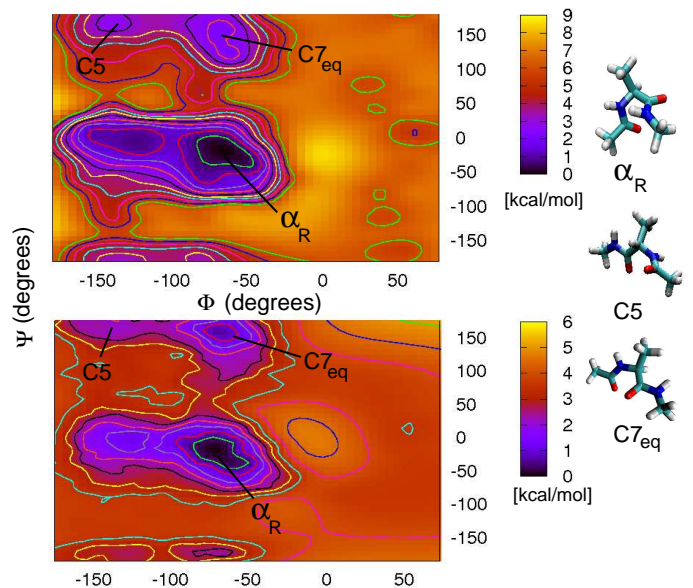

FIG. 7: **Top:** Positions in the eigenvector space in the 300 K biased simulation (orange triangles) and the unperturbed simulation results which are illustrated as blue stars. **Bottom:** Corresponding values for the dihedral angles Φ and Ψ sampled from the same simulation runs.

FIG. 9: **Top:** Free energy landscape derived by the projection of the eigenvectors 1 and 2 on the phase space of the dihedral angles Φ and Ψ. The marked regions denote stable configurations with the corresponding molecular representation. The lines correspond to energy differences of 0.5 kcal/mol. **Bottom:** Corresponding free energy landscape derived by a direct metadynamics simulation in the phase space of the corresponding dihedral angles Φ and Ψ.

unlikely minima to accelerate the rare events of these structural transitions.

The final free energy landscapes at $T = 300$ K are shown in Figs. 8 and 9. Fig. 8 presents the free energy landscape in the space of eigenvectors whereas Fig. 9 is the corresponding Ramachandran plot in the space of the dihedral angles which has been derived by the proposed projection scheme and a direct conventional metadynam-
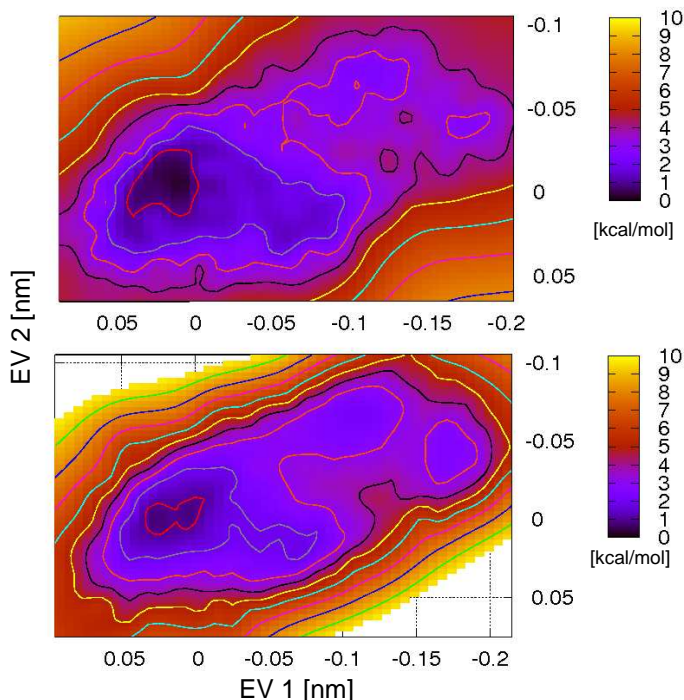
FIG. 10: Free energy landscape for the first two eigenvectors in the conventional metadynamics scheme (top) and in the histogram reweighted metadynamics scheme (bottom). The lines denote energy differences of 0.5 kcal/mol.
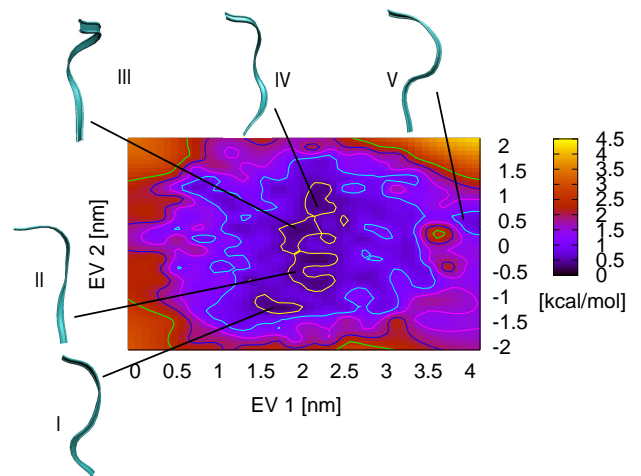

FIG. 11: Free energy landscape for the eigenvectors 1 and 2 of the Met-Enkephalin. The lines correspond to energy differences of 0.5 kcal/mol.

ics simulation run in the phase space of the dihedral angles. Both figures illustrate the corresponding conformations of the alanine dipeptide and their relative energy difference from the most stable configuration α_R to two further minima $C7_{eq}$ and $C5$. The relative energy difference betweeen the α_R and the $C7_{eq}$ and $C5$ minima is given by $\Delta F \approx 1.9$ kcal/mol, respectively 2.2 kcal/mol which is in good agreement to the results reported in Ref. [63, 64] with ≈ 1.74 kcal/mol, respectively ≈ 2.10 kcal/mol. The good agreement in the location and the acceptable values of the free energy minima in comparison to a direct biasing between the plots shown in Fig. 9 are remarkable which validates the proposed projection scheme. Finally we compare the results of the histogram reweighted metadynamics simulation in the phase space of the eigenvectors to a conventional metadynamics simulation with 2000 hills. As the landscapes are nearly identical, it can be concluded that our method is valid. Only slight deviations can be observed for the specific shape of the minina due to the statistical error inherent in the original metadynamics variant [27].

Met-Enkephalin

The Met-Enkephalin has attracted broad interest in its stable conformations [81–84] due to its biological impor-

tance. Previous studies have shown that the main shape of the free energy landscape is given by a funnel [83–85] with a couple of stable conformations which are separated by low energy barriers. Regarding the biological function of the Met-Enkephalin, which is an opioid peptide that inhibits neurotransmitter release from the appropriate opioid receptor, several receptors have to bind on the molecule which all require different stable conformations [86, 87].

The corresponding eigenvector analysis of the Met-Enkephalin illustrates that over 73% of the atomic positional fluctuation have been captured by the first two eigenvectors. Thus nearly all relevant structural transitions can be described by eigenvectors 1 and 2.

The corresponding free energy landscape at $T = 300$ K is shown in Fig. 11. A couple of stable minima and configurations can be found in the funnel-like landscape in agreement to other publications [83–85]. The lines correspond to energy differences of 0.5 kcal/mol. It is obvious that all energetic barriers are lower than 1.5 kcal/mol. Hence the transitions between the stable configurations are not drastically hindered. Regarding the conformations shown in Fig. 11, differences in their form and shape can be observed reflecting the biological function of the Met-Enkephalin [87]. The energetic flexibility in the illustrated conformations finally explains the lack of a poor experimental convergence to a single structure [88].

Another interesting quantity is the solvent accessible surface area which allows the investigation of the importance of the solvent interactions for the Met-Enkephalin. The influence of the hydrophilic solvent accessible surface area A_s^+ , which represents the influence of hydration effects on the stability of the peptide can be investigated by the ratio to the total solvent accessible surface area A_s^t which is illustrated in Fig. 12.
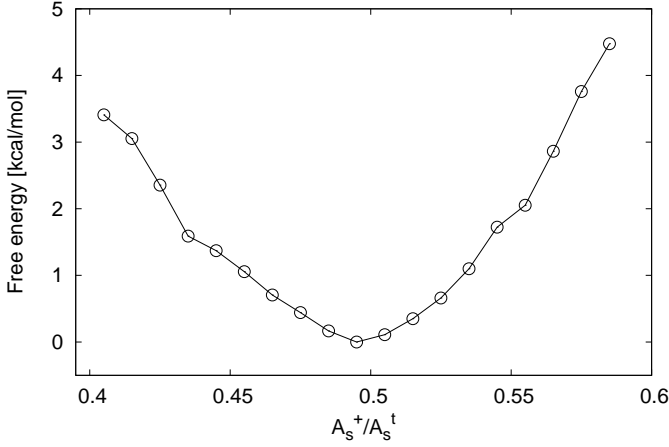
FIG. 12: Free energy landscape for the ratio of the hydrophilic solvent accessible surface area A_s^+ to the total solvent accessible surface area A_s^t.

A large minimum can be found at a ratio of $A_s^+/A_s^t \approx$ 0.49. It can be shown that nearly all stable configurations presented in Fig. 11 obey this characteristic value for the ratio. Hence the formation of different stable conformations can be partly explained by hydration effects due to the chemical structure and solubility of the Met-Enkephalin. The rapid increase in the free energy of 4 kcal/mol for fluctuations $\delta(A_s^+/A_s^t) \approx 0.1$ avoids the appearance of drastic structure transitions around this minimum.

The rigidity and flexibility of the Met-Enkephalin and its residues is finally illustrated in Fig. 13 by a Ramachandran plot. Here the residues GLY-2, GLY-3 and PHE-4 are investigated concerning their mechanical flexibility. All residues are able to form β-sheets while especially the dihedral angles of PHE-4 prefer to visit the α-helical conformations [83, 89, 90]. Left-handed helical conformations can be found in GLY-2 and GLY-3.

COMPUTATIONAL COST

We finally have investigated the computational cost of the histogram reweighted metadynamics scheme in comparison to the ordinary metadynamics technique studied by the two dimensional test case of Eqn. (16). Thus we simulated the two dimensional test case by a conventional metadynamics scheme [27] where every hundredth step a gaussian hill of height 0.1ϵ with a width of 0.25σ has been set. The same values have also been used in the grid scheme. We compared the times needed for the construction of the potential energy landscape using the x and y direction as reaction coordinates and normalized them.

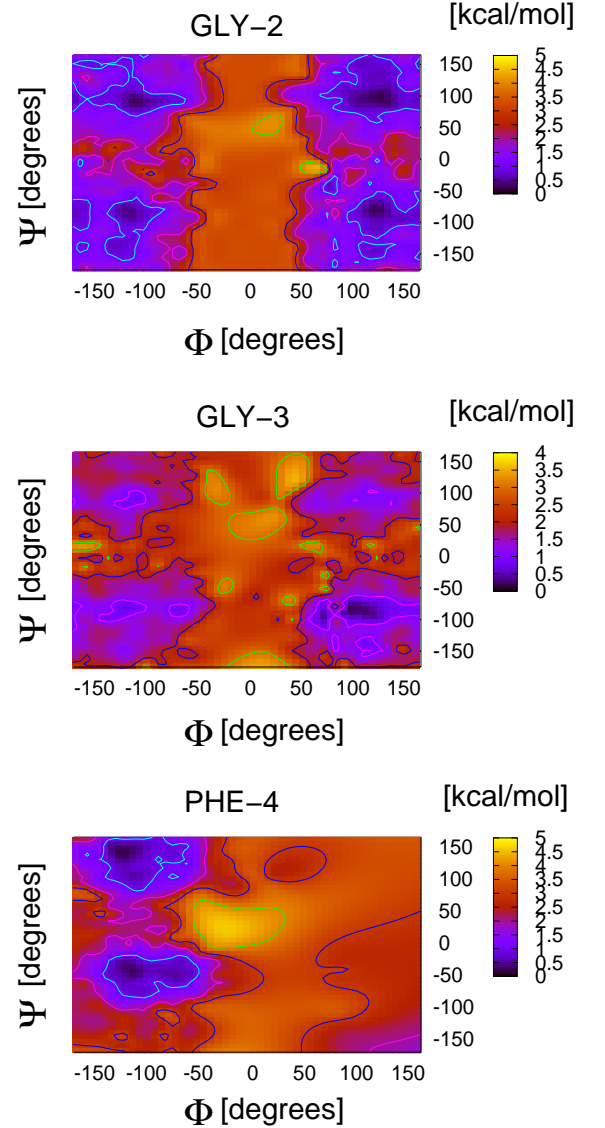To analyze the influence of the number of collective vari-

FIG. 13: Free energy landscape in the Ramachandran plot for three residues. The lines correspond to energy differences of 1.0 kcal/mol.

ables d we conducted simulations with $1, 2, 3, 5$ and 10 collective variables. It is obvious that the grid metadynamics scheme scales as $\mathcal{O}(dt)$ whereas conventional metadynamics obeys a $\mathcal{O}(dt^2)$ behaviour (Fig. 14). Although the program is not computationally optimized due to output-input data flow, it becomes clear that the overall time needed for the conventional metadynamics algorithm is increasing with the second power of simulation steps, respectively present hills. Hence it is evident that the number of hills is the dominating factor. This
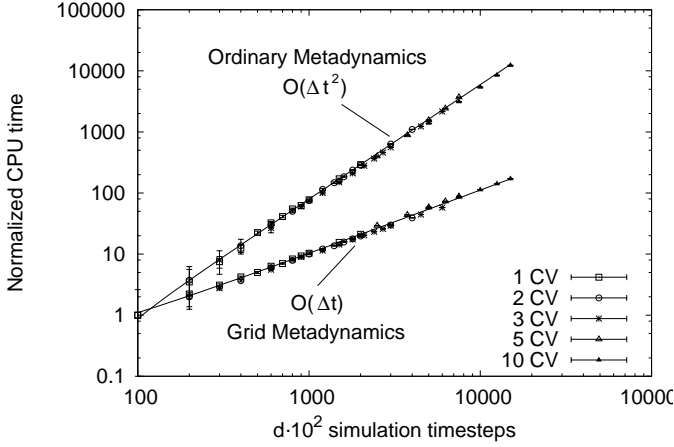
FIG. 14: Normalized computational cost for the construction of the potential energy landscape of Eqn. (16) in the grid scheme in comparison to conventional metadynamics for different numbers of collective variables $d = 1, 2, 3, 5$ and 10.

FIG. 16: Time needed for computation of the ordinary metadynamics forces τ_{MTD} of Met-Enkephalin and alanine dipeptide in comparison to the time for unbiased simulations τ_{ub} for a varying number of hills n_H and number of relaxation steps N_r. All values follow a linear relation with slope $c \sim 0.13$ and the number of collective variables was $d = 2$.
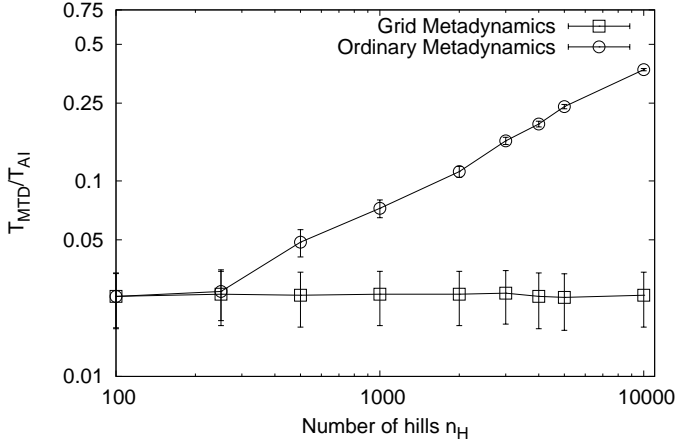

FIG. 15: Ratio for the time needed to calculate the metadynamics forces versus all interactions for one timestep in dependence of an increasing number of hills for the alanine dipeptide.

leads us to the conclusion that our method can be used for a larger number of dimensions as well with a better scaling behaviour than the conventional scheme.

The ratio of the time used for the calculation of the metadynamics algorithm T_{MTD} compared to the time used for all interactions T_{AI} for an increasing number of apparent hills n_H is shown in Fig. 15. All data points have been derived by the simulation of the alanine dipeptide. It is evident that after 250 hills the calculational cost increases drastically in the ordinary metadynamics algorithm whereas for the grid technique the required time remains constant. Thus the grid variant of the metadynamics algorithm accelerates the computation of free energy landscapes in contrast to ordinary metadynamics
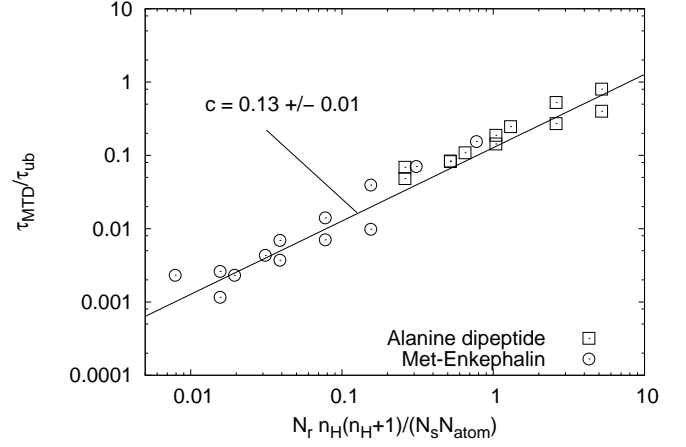
considerably.

Nevertheless it can be assumed that the total time needed for the evaluation of the biasing forces in one timestep is negligible compared to the number of interactions especially for systems with many degrees of freedom. Thus for a system with a large number of interactions, the cost for the evaluation of the metadynamics forces may decrease in comparison to the total time.

To investigate this situation and to derive an empirical formula, we studied the computational cost for the Met-Enkephalin and the alanine dipeptide. Hence we compared the total time which is required for the computation of the conventional metadynamics forces without a grid τ_{MTD} to the total time of an unbiased run τ_{ub}. It can be assumed that the time for the calculation of the metadynamics forces depends on the number of present hills n_H, the number of collective variables d and the number of relaxation steps N_r after that a new hill is deposited. In addition, the contributing number of atoms in the collective variable may also influence the computation time. For reasons of simplicity we assume that this time is mainly determined by the evaluation of $\mathcal{G}(x - x_{G,i})$ (Eqn. (2)) where all other effects are negligible. Thus for much larger systems this factor may become size-dependent.

In summary the total time needed for the evaluation of the ordinary metadynamics forces in a simulation can be written as $\tau_{MTD} \sim dN_r \sum_{j=1}^{n_H} j \sim dN_r n_H n_H (n_H + 1)\delta t_{MTD}/2$ where δt_{MTD} denotes the time needed for the evaluation of a single hill. Additionally it is further assumed that the simulation has finished after n_H hills have been set which is given by the relation $n_H = N_s/N_r$. Furthermore the total time for an unbiased simulation τ_{ub}

is nearly proportional to the number of atoms in the system N_{atom} [65] and the number of simulation timesteps N_s which yields $\tau_{ub} \sim N_{atom} N_s \delta t_{ub}$ with the average time δt_{ub} needed for the calculation of a single unbiased force. This gives the ratio

$$\frac{\tau_{MTD}}{\tau_{ub}} \sim \frac{d N_r n_H (n_H + 1)}{2 N_s N_{atom}} \frac{\delta t_{MTD}}{\delta t_{ub}} \sim \frac{d n_H}{2 N_{atom}} \frac{\delta t_{MTD}}{\delta t_{ub}} \quad (20)$$

where $\delta t_{MTD}/\delta t_{ub}$ can be empirically determined. It has to be mentioned that the ratio is strongly dependent on the algorithms which are used for the evaluation of the forces but not on the processor speed.

We have used varying relaxation timesteps of $N_r = 1 - 1000$, different numbers of hills $n_H = 50 - 10000$ and the number of atoms was $N_{atom} = 3231$ for the Met-Enkephalin with water and $N_{atom} = 961$ for the alanine dipeptide plus water whereas N_s varies between 100 and 100000.

The values for different ordinary metadynamics runs in comparison to unbiased runs for the Met-Enkephalin and the alanine dipeptide are shown in Fig. 16. It is evident that all values follow Eqn. (20) with a proportionality factor of $c = 0.13 \pm 0.01$. Hence for a large number of present hills, the ordinary metadynamics method drastically increases computation time. Thus as a general remark, the original method becomes computationally very expensive if the number of hills is larger than the number of atoms. This is often given for systems with implicit solvent models or complex and large energy landscapes. As we have shown before, the grid based technique obeys a linear $\mathcal{O}(dt)$ behaviour, such that the ratio between the time needed for an ordinary metadynamics algorithm in comparison to the grid technique also grows in accordance to Eqn. (20). Especially for ordinary metadynamics simulations where the free energy landscape is not refined by histogram reweighting procedures the amount of settled hills often increases the number of atoms which strongly suggests the use of grid based techniques. This is mostly important for classical Molecular Dynamics simulations for which our method is more intended in contrast to ab-intio methods. The evaluation of the electronic motion is here the main factor [92] dominating most of the computation time such that the evaluation of the metadynamics potential can be neglected.

SUMMARY AND CONCLUSION

We have presented an efficient and simple method for the calculation of free energy landscapes. The technique is applicable for a broad range of different systems. The basic principles are the construction of a potential landscape on a predefined grid in an initial simulation run which is used as a biasing potential in the final simulations. The corresponding probability distributions of the biased simulations are reweighted by the WHAM procedure whose usage avoids specific drawbacks of the ordinary metadynamics algorithm like large error tolerances [46].

Another advantage is the easy implementation due to its simple methodology in software packages like GROMACS. In addition we have presented the application of a projection scheme which allows to transform the energy landscape of certain collective variables to further reaction coordinates without any additional simulation effort. The specific choice of biasing the first eigenvectors of the system achieves a good activation of the whole system. The history-dependent potential is short-ranged in contrast to the ordinary Metadynamics scheme [26, 27] and serves as a pure biasing potential. The fine resolution of the free energy landscape is achieved by histogram reweighting techniques of the corresponding probability distributions. A further advantage of the method is the opportunity to tune the resolution of the landscape even after the simulations were finished.

Furthermore we have explicitly shown that a grid-based technique scales with $\mathcal{O}(dt)$ in contrast to the conventional metadynamics scheme ($\mathcal{O}(dt^2)$), where t denotes the simulation time and d the number of applied collective variables. Hence, compared to the ordinary metadynamics algorithm a grid technique is more preferable due to a decreased simulation time. This has been shown by the computational cost for the alanine dipeptide and the Met-Enkephalin. Via a linear relation we were able to show that the computational cost of the ordinary metadynamics method in comparison to the grid technique can be predicted by an empirical formula. Thus the grid method is preferable for systems where the number of settled hills exceeds the number of all atoms in the system. As we have discussed, this fact is given for implicit solvation models as well as very large energy landscapes. In addition our method has been tested for the peptides alanine dipeptide and Met-Enkephalin. The results are in good agreement to the literature. We have further shown that the energy landscape of the Met-Enkephalin is funnel-like with many different conformations at several energetic minima. The similarity between these conformations can be illustrated by a characteristic ratio of the hydrophilic solvent accessible surface area in comparison to the total solvent accessible area which results in a characteristic minimum around $A_s^+/A_s^t \approx 0.49$. The flexibility of the Met-Enkephalin is illustrated by plotting the dihedral angles of each residue in a Ramachandran plot.

Additionally we have shown that the method allows to identify the stable conformations of the alanine dipeptide with good accuracy of the free energy landscape. Following [27, 46], where it was recognized that the functionality of a biasing potential does not depend on its specific shape, we have further validated our technique by the explicit calculation of two well-defined test cases

such that it can be used as an additional variation of the existing metadynamics methods.

ACKNOWLEDGEMENTS

---

* Electronic address: jens.smiatek@uni-muenster.de
† Electronic address: andheuer@uni-muenster.de

[1] Dobson C. M., Sali A. and Karplus M., *Angew. Chem.* **110**, 908 (1998)
[2] Onuchic J. N., Schulten, Z. L. and Wolynes, P. G., *Annu. Rev. Phys. Chem.* **48**, 545 (1997)
[3] Wales D. J. and Bogdan, T. V., *J. Chem. Phys. B* **110**, 20765 (2006)
[4] Wales D. J., *Int. Rev. Phys. Chem.* **25**, 237 (2006)
[5] Dill K. A., Banu Ozkan, S., Scott Shell M. and Weikl, T. R., *Annu. Rev. Biophys.* **37**, 289, (2008)
[6] Heuer A., *J. Phys.: Cond. Mat.* **20**, 373101 (2008)
[7] Wales D. J., Miller M. A. and Walsh T. R., *Nature* **394**, 758 (1998)
[8] Wales, D. J., *Energy Landscapes.* Cambridge: Cambridge University Press (2003)
[9] Bryngelson, J. D., Onuchic, J. N., Socci, N. D. and Wolynes, P.G., *Proteins* **21**, 167 (1995)
[10] Wales D. J., *Mol. Phys.* **100**, 3285 (2002)
[11] Dellago C., Bolhuis P. G., Csajka F. S. and Chandler D., *J. Chem. Phys.* **108**, 1964 (1998)
[12] Dellago C., Bolhuis P. G. and Geissler P. L., *Adv. Chem. Phys.* **123**, 1 (2002)
[13] Kollman P., *Chem. Rev.* **93**, 2395 (1993)
[14] Carter E. A., Ciccotti G., Hynes J. T. and Kapral R., *Chem. Phys. Lett.* **156**, 472 (1989)
[15] Sprik M. and Cicotti G., *J. Chem. Phys.* **109**, 7737 (1989)
[16] Bash P. A., Singh U. C., Brown F. K., Langridge P. and Kollman P. A., *Science* 235, 574 (1987)
[17] Patey G. N. and Valleau J. P., *J. Chem. Phys* **63**, 2334 (1975)
[18] Darve E. and Pohorille A., *J. Chem. Phys.* **115**, 9169 (2001)
[19] Darve E., Rodriguez-Gomey D. and Pohorille A., *J. Chem. Phys.* **128**, 144120 (2008)
[20] Gulligsrud J., Braun R. and Schulten K., *J. Comput. Phys.* **151**, 190 (1999)
[21] Rosso L., Minary P., Zhu Z. and Tuckerman M., *J. Chem. Phys.* **116**, 4389 (2002)
[22] Berg B. A. and Neuhaus T., *Phys. Rev. Lett.* **267**, 249 (1991)
[23] Berg B. A. and Neuhaus T., *Phys. Rev. Lett.* **68**, 9 (1992)
[24] Mitsutake A., Sugita Y. and Okamoto Y. *Biopolymers* **60**, 96 (2001)
[25] Earl D. J. and Deem, M. W., *Phys. Chem. Chem. Phys.* **7**, 3910 (2005)
[26] Laio A. and Parrinello M., *Proc. Natl. Acad. Sci. USA* **99**, 12562 (2002)
[27] Laio A. and Gervasio F. L., *Rep. Prog. Phys.* **71**, 126601 (2008)
[28] Huber T., Torda A. E. and van Gunsteren, W. F., *J. of Computer-Aided Molecular Design* **8**, 695 (1994)
[29] Grubmüller H., *Phys. Rev. E* **52**, 2893 (1995)
[30] Tiana G., *Europ. Phys. J. B* **63**, 235 (2008)
[31] Laio A., Rodriguez-Fortea A., Gervasio F. L., Ceccarelli M. and Parrinello M., *J. Phys. Chem. B* **109**, 6714 (2005)
[32] Bussi G., Laio A. and Parrinello M., *Phys. Rev. Lett.* **96**, 090601 (2006)
[33] Barducci A., Bussi G. and Parrinello, M., *Phys. Rev. Lett.* **100**, 020603 (2008)
[34] Bonomi M., Barducci A. and Parrinello M., *J. Comp. Chem.* **30**, 1615 (2009)
[35] Barducci A., Bonomi M. and Parrinello M., *Biophys. J.* **96**, 44 (2010)
[36] Micheletti C., Laio A. and Parrinello M., *Phys. Rev. Lett.* **92**, 170601 (2004)
[37] Ferrenberg A. and Swendsen R., *Phys. Rev. Lett.* **61**, 2635 (1988)
[38] Kumar S., Rosenberg J. M., Bouzida D., Swenden R. and Kollman P. A., *J. Comput. Chem.* **16**, 1339 (1992)
[39] Roux B., *Comput. Phys. Comm.* **91**, 275 (1995)
[40] Chodera J. D., Swope W. C., Pitera J. W., Seok C. and Dill K. A., *J. Chem. Theory Comput.* **3**, 26 (2007)
[41] Krivov S. V. and Karplus M., *PNAS 101*, 14766 (2004)
[42] Muff S. and Caflisch A., *Proteins* **70**, 1185 (2008)
[43] Noe F. and Fischer S., *Curr. Op. Struct. Biol.* **18**, 154 (2008)
[44] Strodel B. and Wales D. J., *Chem. Phys. Lett.* **466**, 105 (2008)
[45] Ensing B., Laio A., Gervasio F. L., Parrinello M., and Klein M. L., *J. Am. Chem. Soc.* **126**, 9492 (2004)
[46] Babin V., Roland C., Darden T. A. and Sagui C., *J. Chem. Phys.* **125**, 204909 (2006)
[47] Babin V., Roland C. and Sagui C., *J. Chem. Phys.* **128**, 134101 (2008)
[48] Marinelli F., Pietrucci F., Laio A. and Piana S., *PLOS* **5**, 1000452 (2009)
[49] Bartels C. and Karplus M., *J. Comp. Chem.* **18**, 1450 (1997)
[50] Bonomi M., Branduardi D., Bussi G., Camilloni C., Provasi D., Raiteri P., Donadio D., Marinelli F., Pietrucci F., Broglia R. A. and Parrinello M., *Comp. Phys. Comm.* **180**, 1961 (2009)
[51] Plumed-Manual Version 1.2 (September 2010)
[52] Henin J., Fiorin G., Chipot C. and Klein M. L., *J. Chem. Theory Comput.* **6**, 35 (2010)
[53] NAMD User's Guide: *http://www.ks.uiuc.edu/Research/namd/2.7b3/ug/* (September 2010)
[54] Case D. A., Darden T. A., Cheatham T. E., Simmerling C. L., Wanj J., Duke R. E., Luo R., Walker R. C., Zhang W., Merz K. M., Roberts B., Wang B., Hayik S., Roitberg A., Seabra G., Kolossvary I., Wong K. F., Paesani F., Vanicek J., Liu J., Wu X., Brozell S. R., Steinbrecher T., Gohlke H., Cai Q., Ye X., Wang J., Hsieh M.-J., Cui G., Roe D. R., Mathews D. R., Seetin M. G., Sagui C., Babin V., Luchko T., Gusarov S., Kovalenko A. and Kollman P. A., *AMBER 11*, University of California, San Francisco

[55] Case D. A., Cheatham T., Darden T., Gohlke H., Luo R., Merz Jr. K. M., Onufriev A., Simmerling C., Wang B. and Woods R., *J. Comput. Chem.* **26**, 1668 (2005)

[56] Wang F. and Landau D., *Phys. Rev. Lett.* **86**, 2050 (2001)

[57] Babin V. and Sagui C., *J. Chem. Phys.* **132**, 104108 (2010)

[58] Berendsen H. J. C., van der Spoel D., van Drunen R., *Comp. Phys. Comm.* **91**, 43 (1995)

[59] Hess B., Kutzner C., van der Spoel D. and Lindahl E., *J. Chem. Theory Comput.* **4**, 435 (2008)

[60] Van der Spoel D., Lindahl E., Hess B., Groenhof, G., Mark A. E. and Berendsen H. J. C., *J. Comp. Chem.* **26**, 1701 (2005)

[61] Raiteri P., Laio A., Gervasio F. L., Micheletti C. and Parrinello M., *J. Phys. Chem. B* **110**, 3533 (2006)

[62] Amadei A., Linssen A. B. and Berendsen H. J. C., *Proteins* **17**, 412 (1993)

[63] Spiwok V., Lipova P. and Kralova B., *J. Phys. Chem. B* **111**, 3073 (2007)

[64] Spiwok V., Kralova B. and Tvaroska I., *J. Mol. Model* **14**, 995 (2008)

[65] Frenkel D. and Smit B., *Understanding Molecular Simulation*, Academic Press, San Diego (1996)

[66] Laio A., Rodriguez-Fortea A., Gervasio F. L., Ceccarelli M. and Parrinello M., *J. Phys. Chem. B* **109**, 6714 (2005)

[67] Schlick T., *Molecular Modelling and Simulation*, Springer Press, New York (2002)

[68] Cornell W. D., Cieplak P., Bayly C. I., Gould I. R., Merz K., M., Ferguson D. M., Spellmeyer D. C., Fox T., Caldwell J. W. and Kollman P. A., *J. Am. Chem. Soc.* **117** , 5179 (1995)

[69] Sorin E. J., Rhee Y. M. and Pande V. S., *Biophys. J.* **88**, 2472 (2005)

[70] Jorgensen W. L., Chandrasekhar J., Madura D. J., Impey R. W. and Klein M. L. *J. Chem. Phys.* **79**, 926 (1983)

[71] Ryckaert J.-P., Ciccotti G. and Berendsen H.J.C., *J. Comp. Phys.* **23**, 327 (1977)

[72] Marcotte I., Separovic F., Auger, M. and Gagne, S.M., *Biophys. J.* **86**, 1587 (2004)

[73] Oostenbrink C., Villa A., Mark A. E. and Van Gunsteren W. F., J. Comp. Chem. **25**, 1656 (2004)

[74] Smith P., *J. Chem. Phys.*, **111**, 5568 (1999)

[75] Gfeller D., De Los Rios P., Caflisch A. and Rao F., *PNAS USA* **104**, 1817 (2007)

[76] Chekmarev D. S., Ishida T. and Levy R. M., *J. Phys. Chem. B* **108**, 19487 (2004)

[77] Ren W., Vanden-Eijnden E., Maragakis P. and Weinan E., *J. Chem. Phys.* **123**, 134109 (2005)

[78] Ma A. and Dinner A. R., *J. Phys. Chem. B* **109**, 6769 (2005)

[79] Branduardi D., Gervasio F. L. and Parrinello M., *J. Chem. Phys.* **126**, 054103 (2007)

[80] Pan A. C., Sezer D. and Roux B., *J. Phys. Chem. B* **112**, 3432 (2008)

[81] Koca J. and Carlsen H. J., *J. Mol. Struct.* **337**, 17 (1995)

[82] Perez J. J., Loew G. H. and Villar H. O., *Int. J. Quantum Chem.* **44**, 263 (1992)

[83] Sanbonmatsu K. Y. and Garcia A. E. *Proteins* **46**, 225 (2002)

[84] Hansmann U. H. E., Okamoto Y. and Onuchic J. N. *Proteins* **34**, 472 (1999)

[85] Evans D. A. and Wales D. J., *J. Chem. Phys.* **119**, 9947 (2003)

[86] Hughes J., Smiath T. W., Kosterlitz H. W., Fothergill L. A., Morgan B. A. and Morris H. R. *Nature* **258**, 577 (1975)

[87] Kendrew J. C. *The encyclopedia of molecular biology.* Oxford: Blackwell Science (1994)

[88] Graham W. H., Carter E. S. II and Hicks R. P. *Biopolymers* **32**, 1755 (1992)

[89] Cornell W. D., Caldwell J. D. and Kollman P. A., *J. de Chimie Phys. et de Physico-Chimie Biol.* **94**, 1417 (1997)

[90] Head-Gordon T., Head-Gordon M., Frisch M., Pople J. A. and Brooks C. L., *J. Am. Chem. Soc.* **113**, 5989 (1991)

[91] *http://web.vscht.cz/spiwokv/index.html* (2010)

[92] Leach A., *Molecular Modeling: Principles and Applications.* Prentice-Hall: New York (2001)